

Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics

John J. Portman^{a)}

Department of Physics, University of Illinois, Urbana, Illinois 61801

Shoji Takada

Department of Chemistry, Kobe University, Kobe 657-8501, Japan

Peter G. Wolynes^{b)}

Department of Chemistry, University of Illinois, Urbana, Illinois 61801

(Received 31 August 2000; accepted 30 October 2000)

The motions involved in barrier crossing for protein folding are investigated in terms of the chain dynamics of the polymer backbone, completing the microscopic description of protein folding presented in the preceding paper. Local reaction coordinates are identified as collective growth modes of the unstable fluctuations about the saddle points in the free energy surface. The description of the chain dynamics incorporates internal friction (independent of the solvent viscosity) arising from the elementary isomerization of the backbone dihedral angles. We find that the folding rate depends linearly on the solvent friction for high viscosity, but saturates at low viscosity because of internal friction. For λ -repressor, the calculated folding rate prefactor, along with the free energy barrier from the variational theory, gives a folding rate that agrees well with the experimentally determined rate under highly stabilizing conditions, but the theory predicts too large a folding rate at the transition midpoint. This discrepancy obtained using a fairly complete quantitative theory inspires a new set of questions about chain dynamics, specifically detailed motions in individual contact formation. © 2001 American Institute of Physics. [DOI: 10.1063/1.1334663]

I. INTRODUCTION

A complete description of protein folding kinetics must account for the microscopic dynamics of the polypeptide chain. The relaxation of a free polymer is well described by Langevin dynamics with an harmonic potential between adjacent monomers and Markovian random forces.¹ This Rouse–Zimm formalism can be extended through frictional forces with memory to effectively describe chain motions that include many other strictly microscopic complexities such as chain stiffness.^{2,3} To describe the chain motions involved in protein folding, one must also attend to the residue specific interactions or free energy landscape which guides the protein to its native state and governs the probability of different ensembles of structures described in terms of order parameters.

For many proteins, the only thermodynamically distinct states with significant detectable populations are the unfolded ensemble and the native folded conformations.⁴ This observation suggests the existence of a free energy barrier separating the stable and metastable phases. Like nucleation in phase transitions, the folding kinetics for these proteins is largely governed by the probability of rare configurations corresponding to the top of the barrier, the so-called transition state ensemble. In theoretical studies, the transition state

ensembles correspond with saddle points in a multidimensional free energy surface as a function of local order parameters. Near the transition state, some order parameter fluctuations grow to become the folded phase. This is an unstable mode of collective motion. Motion in other directions tend to stabilize the transition state. The structure of the unstable growth mode can be thought of as a locally defined reaction coordinate. If there are multiple saddle points which are connected, the local reaction coordinate for each of these will usually differ. In such cases, especially, the local reaction coordinate may not be the best way of describing the global transformation of the ensemble involved in folding.

Lately, there has been a great effort to elucidate the nature of the transition state ensemble and ideal reaction coordinate for specific proteins through measurement,^{5–10} theoretical calculations based on physically intuitive choices,^{11–15} and simulations.^{16–22} While many aspects of the folding kinetics seem to be understood, still some basic issues remain controversial. For example, while the structural distribution of the transition state ensemble is well predicted,^{11–15} the absolute magnitude of the barrier height is not well determined. Comparison with experiment is clouded by the uncertainty of the prefactor of the folding rate.

For simple thermally activated transitions, one expects the rate to follow the activated law

$$k = k_0 e^{-\beta \Delta F^\ddagger}, \quad (1)$$

where $\beta = 1/k_B T$ is the inverse temperature and ΔF^\ddagger is the

^{a)}Current address: Theoretical Biology and Biophysics Group, T-10 MSK710, Los Alamos National Laboratory, Los Alamos, New Mexico 87545.

^{b)}Current address: Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, California 92093.

barrier height which contains both entropy and energy. For folding, ΔF^\ddagger is expected to be strongly temperature dependent. The exponential factor is the relative probability of finding the system at the transition state compared to the metastable minimum. The prefactor, k_0 , is determined by the microscopic dynamics involved in crossing the barrier. For proteins with very rugged landscapes, the dynamics can be described as hopping between trapped states.²³ In this case the barriers to escape traps enter strongly. For proteins with very little frustration that have smoother landscapes, there will be little trapping, but the chain motion will be diffusive and depend on chain connectivity. k_0 depends on the free energy surface and the frictional forces such as viscosity that determine the stochastic motion of the protein as it folds. One estimate proposed for the prefactor relates k_0 to the inverse time it takes to form a typical loop given by simple polymer theory, $k_0 \approx 1 \mu\text{s}^{-1}$.^{24,25} While quite reasonable, this estimate does not involve the shape of the free energy surface near the transition state. Thus, for example, it is not completely clear how big a loop should be considered.

In Ref. 12 and the preceding paper (henceforth referred as paper I), we presented a variational theory for the protein folding free energy surface of a specific protein to a known structure. In this approach, the protein is modeled as a collapsed, stiff chain with interactions between residues in contact in the native structure. The reference Hamiltonian is that of a polymer chain inhomogeneously constrained to the native structure. The free energy profile for folding obtained from the model can be described as an activation barrier between the unfolded and folded ensembles with fine structure from local minima and saddle points in the free energy surface giving rise to longitudinal ruggedness in the free energy profile for a perfectly funneled landscape. In this approach, conformational ensembles along the folding route are characterized by structural order parameters evaluated as averages over the distribution from the reference Hamiltonian. The focus of the present paper is how the microscopic chain dynamics are involved in barrier crossing within this model.

Understanding barrier crossing dynamics in protein folding requires not only a good characterization of the free energy surface and structural ensembles, but also a good description of the microscopic motions of the polymer backbone. In the Rouse–Zimm model of polymer dynamics, the motion of the polymer chain is governed by the chain connectivity, represented as harmonic entropic springs connecting adjacent monomers, and is damped by the interactions of the monomers with the solvent.¹ A more detailed description of polypeptide dynamics acknowledges there are activated transitions between *gauche* and *trans* bond angles determined by the dihedral potential.^{26–28} One way to incorporate the short range structural information arising from these barriers to internal rotation is through a static chain stiffness that defines renormalized harmonic correlations.^{3,29–36} However, a complete description must also incorporate the activated nature of these rotational isomerization rates in the dynamical correlations.

Quite early in the study of polymer dynamics, anomalies found in intrinsic viscosity measurements indicated that there

was a source of friction independent of the viscosity of the solvent.^{37,38} Kuhn and Kuhn first suggested this “internal viscosity” was a consequence of microscopic barrier hopping in the local chain segments.³⁷ This phenomenological friction found for high polymers may have a variety of microscopic origins in addition to internal friction due to barrier hopping of many elementary chain links. A helpful (though perhaps dated) evaluation of different theories can be found in Ref. 39. Fixman has shown that simulations for polymers with rigidly fixed bond angles and bond lengths can be mimicked by internal viscosity.^{40,41} Still other possible sources of the measured internal viscosity include the effect of interactions with distant monomers in the chain,⁴² or other inherent nonlinearities neglected in the Gaussian model assumption.⁴³ To our knowledge, the microscopic interpretation of internal viscosity for artificial high polymers has not been definitively settled. It is clear, however, that polypeptides will at least exhibit internal viscosity from the difficulty of making backbone angle transitions. These transitions are roughly 100 times slower than if no barriers were present in the Ramachandran potential.

In this paper, we complete the logical development of folding rate theory for minimally frustrated proteins by analyzing the microscopic dynamics of chain motions involved in barrier crossing of the protein folding model presented in Ref. 12 and paper I. We use multidimensional, non-Markovian Langevin dynamics to describe the barrier crossing, and identify the unstable modes as the local reaction coordinates. For the folding rate calculation, we use a generalization of Kramer’s theory⁴⁴ that treats non-Markovian dynamics in many dimensions.⁴⁵ At the same time, the interpretation of a local reaction coordinate follows Langer’s theory of the nucleation rates.⁴⁶ The details of chain dynamics enter the formalism through a generalized friction matrix with memory. The friction that damps the motion of the chain includes the solvent viscosity and an internal friction that depends on the length scale (in sequence) of the motion but which we take as independent of the solvent viscosity. The internal friction accounts for the activated nature of dihedral angle flips in polypeptides. We restrict our attention to overdamped dynamics, though it is easy to generalize the calculations to include inertial terms as well.

II. REACTION COORDINATES AND FOLDING RATES

To calculate the barrier crossing rate, one must first identify the order parameters that describe the phases of the system, and then study their motions. In density functional theory, one often considers a density field that is slowly varying in space, and expands the density functional about a uniform density.⁴⁷ This expansion yields a free energy consisting of a bulk free energy and can be expanded in gradient terms that ultimately represent the surface tension between two phases. The density then evolves either hydrodynamically or diffusively under the influence of the free energy driving force. The quasiequilibrium part of this approach has been applied to nucleation of the folded state in proteins starting with a mean field free energy functional and defining a global native density field as a measure of similarity to the native state⁴⁸

$$\rho^N(\mathbf{r}, \{\mathbf{r}_{ij}\}) \sim \sum_i \delta(\mathbf{r} - \mathbf{r}_i) \delta(\mathbf{r}_i - \mathbf{r}_i^N), \quad (2)$$

where $\{\mathbf{r}_{ij}\}$ specify the configuration of the protein chain, and $\{\mathbf{r}_i^N\}$ are the corresponding coordinates of the native structure. In the present work, we wish to have a description of the folding that is local in sequence and hence consider each term in the sum separately. This local native density defined as

$$\rho^N(\mathbf{r}, \mathbf{r}_i) = \delta(\mathbf{r} - \mathbf{r}_i) \exp\left[-\frac{3}{2a^2} \alpha^N (\mathbf{r}_i - \mathbf{r}_i^N)^2\right], \quad (3)$$

where we have relaxed the measure of similarity of the native state by replacing $\delta(\mathbf{r}_i - \mathbf{r}_i^N)$ by a Gaussian measure.

To connect this density to our folding route calculations, it is more convenient to consider scalar order parameters than to expand the free energy in terms of field variables. Therefore, we integrate $\rho(\mathbf{r}, \mathbf{r}_i)$ over \mathbf{r} , to give the scalar measure native similarity

$$\rho(\mathbf{r}_i) = \exp\left[-\frac{3}{2a^2} \alpha^N (\mathbf{r}_i - \mathbf{r}_i^N)^2\right]. \quad (4)$$

[To simplify notation, we have suppressed the superscript N since the native density can still be distinguished by the one-body monomer density, $\rho^1(\mathbf{r})$, given in paper I, Eq. (25).] Following the interpretation of the order parameters given in paper I, we calculate the average native density of site i , $\rho_i = \langle \rho(\mathbf{r}_i) \rangle_0$, for a given set of variational constraints $\{C_{ij}\}$,

$$\begin{aligned} \rho_i[\{C\}] &= \int d\mathbf{r} \rho_i^1(\mathbf{r}) \rho(\mathbf{r}) \\ &= (1 + \alpha^N G_{ii})^{-3/2} \exp\left[-\frac{3}{2a^2} \frac{\alpha^N (\mathbf{s}_i - \mathbf{r}_i^N)^2}{1 + \alpha^N G_{ii}}\right]; \end{aligned} \quad (5)$$

that is, ρ_i is a function of $\{C_{ij}\}$ through the correlations, G_{ij} , and average position, \mathbf{s}_i , given by Eqs. (21) and (22) in paper I, respectively. The free energy can then be parameterized by $\{\rho_i\}$ with $\bar{F}[\{\rho\}] = F[\{C\}]$ where $\{\rho\}$ is evaluated at $\{C\}$. In particular, we denote a saddle point of $F[\{C\}]$ by $\{C^*\}$ and the corresponding native density by $\rho_i^* = \rho_i[\{C^*\}]$.

We assume Langevin dynamics with memory for the native density $\{\rho_i(t)\}$,

$$\partial_t \rho_i(t) = - \sum_j \int_0^t dt' \mu_{ij}(t-t') \frac{\partial \beta \bar{F}[\{\rho(t')\}]}{\partial \rho_j} + \xi_i(t), \quad (7)$$

where $\beta = 1/k_B T$ is the inverse temperature, and $\mu(t)$ is a generalized mobility matrix related to the random noise $\xi_i(t)$ through the correlations $\langle \xi_i(t) \xi_j(t') \rangle = \mu_{ij}(t-t')$. The form of $\mu(t)$ is discussed in the following section, here we assume that $\mu(t)$ is a known function of time.

To study the dynamics near the saddle point, we expand $\bar{F}[\{\rho\}]$ about ρ_i^* to second order

$$\beta \bar{F}[\{\rho\}] \approx \text{const} + \frac{1}{2} \sum_{ij} \bar{\Gamma}_{ij}^* \delta \rho_i \delta \rho_j, \quad (8)$$

where $\delta \rho_i = \rho_i - \rho_i^*$, and the Hessian matrix

$$\bar{\Gamma}_{ij}^* = \frac{\partial^2 \beta \bar{F}[\{\rho^*\}]}{\partial \rho_i \partial \rho_j} \quad (9)$$

has one negative eigenvalue since it is evaluated at a saddle point. In terms of derivatives with respect to the variational parameters, $\bar{\Gamma}^*$ is the solution of the matrix equation

$$\sum_{kl} \bar{\Gamma}^{kl} \frac{\partial \rho_k^*}{\partial C_i} \frac{\partial \rho_l^*}{\partial C_j} = \frac{\partial^2 \beta F[\{C^*\}]}{\partial C_i \partial C_j}, \quad (10)$$

where the derivatives with respect to C_i can be calculated as described in paper I.

With this approximation, Eq. (7) can be written as

$$\partial_t (\delta \rho_i(t)) = - \sum_j \int_0^t dt' [\mu(t-t') \bar{\Gamma}^*]_{ij} \delta \rho_j(t') + \xi_i(t). \quad (11)$$

Accordingly, the native density correlation functions $\mathcal{C}(t)$ with components

$$\mathcal{C}_{ij}(t) = \langle \delta \rho_i(t) \delta \rho_j(0) \rangle_0, \quad (12)$$

satisfy

$$\partial_t \mathcal{C}(t) = - \int_0^t dt' \mu(t-t') \bar{\Gamma}^* \mathcal{C}(t'). \quad (13)$$

In Eq. (11), we have subtracted the stationary value $\rho_i(t \rightarrow \infty) = \rho_i^*$, so that $\delta \rho_i(t \rightarrow \infty) = 0$ and $\mathcal{C}(t \rightarrow \infty) = 0$. This long time solution is unstable, however, because it is a saddle point of the free energy.

The unstable mode can be determined by the average equation of motion.⁴⁶ Denoting the Laplace transform of an arbitrary function of time by $\hat{g}(\omega) = \int_0^\infty dt e^{-\omega t} g(t)$, the solution of Eq. (11) averaged over the random noise is

$$\overline{\delta \hat{\rho}_i(\omega)} = [\omega \mathbf{1} + \hat{\mu}(\omega) \bar{\Gamma}^*]_{ij}^{-1} \delta \rho_j(0). \quad (14)$$

By inverting the Laplace transform, $\overline{\rho_i(t)}$ can be expressed as a sum of exponentials with time constants determined by the poles of $[\omega \mathbf{1} + \hat{\mu}(\omega) \bar{\Gamma}^*]^{-1}$: the eigenfunctions $\hat{\mu}(-\kappa) \bar{\Gamma}^* \cdot \mathbf{u} = \kappa \mathbf{u}$ have time dependence $\mathbf{u}(t) \sim e^{-\kappa t}$. Assuming that $\hat{\mu}(\omega)$ is positive definite, there is one mode for which κ is negative, because $\bar{\Gamma}$ is the saddle point Hessian of $F[\rho]$:

$$\mathbf{u}^*(t) = \mathbf{u}^* e^{|\kappa^*|t}, \quad (15)$$

where

$$\hat{\mu}(-\kappa^*) \bar{\Gamma}^* \cdot \mathbf{u}^* = \kappa^* \mathbf{u}^* \quad (\kappa^* < 0). \quad (16)$$

Since motion along the unstable mode \mathbf{u}^* grows exponentially away from the saddle point, we identify the components of \mathbf{u}^* as the local reaction coordinate to surmount the barrier. Both \mathbf{u}^* and κ^* are essentially the unstable mode and curvature of $\bar{F}[\rho]$, but renormalized by the dynamics of the barrier crossing incorporated in the mobility matrix. In Eq. (16), κ^* is the multidimensional analogue of the Grote-Hynes frequency (in the overdamped limit).⁴⁹

The rate for barrier crossing corresponding to Eq. (11) is⁴⁵

$$k = \frac{|\kappa^*|}{2\pi} \left| \frac{\det \bar{\Gamma}_{MS}}{\det \bar{\Gamma}^*} \right|^{1/2} e^{-\beta \Delta F^\ddagger}, \quad (17)$$

where ΔF^\ddagger is the barrier height and $\bar{\Gamma}_{MS}$ and $\bar{\Gamma}^*$ are the curvature matrices of $\bar{F}[\rho]$ evaluated at the metastable minimum and the saddle point, respectively. Equation (17) generalizes both the rate calculations presented by Langer⁴⁶ and Grote–Hynes;⁴⁹ it simultaneously accommodates both multidimensional diffusion (as in the Langer formula) as well as time dependent friction (as in the Grote–Hynes formula).

The ratio of determinants accounts for the entropic differences between the transition state and the metastable phase due to fluctuations of the order parameter. These fluctuations are like the capillary waves which renormalize the surface tension in ordinary nucleation. These contributions would already be included in an exact free energy so that the determinants should be absorbed into the exponential factor to most simply keep a consistent level of thermodynamic theory.⁵⁰ Consequently, the expression for the rate becomes

$$k = \frac{|\kappa^*|}{2\pi} e^{-\beta \Delta F^\ddagger}, \quad (18)$$

where κ^* is given by Eq. (16). As pointed out in Ref. 51, if the overcounting of entropy were not corrected, the ratio of the forward and backwards rate would not equal the equilibrium constant as determined by the starting free energy functional, $k_{12}/k_{21} \neq e^{-\beta(F_2 - F_1)}$.

III. LINEAR RESPONSE APPROXIMATION OF THE GENERALIZED MOBILITY

The time evolution of the native densities $\{\delta\rho_i(t)\}$ depends on microscopic dynamics of the monomer positions $\{\mathbf{r}_i(t)\}$. The chain dynamics, in turn, are incorporated into the effective mobility matrix $\mu(t)$ in the relaxation equation for $\delta\rho(t)$ [Eq. (11)]. To determine $\mu(t)$ consistent with the definition of the native densities this theory, we consider the chain dynamics near the saddle point to be governed by the H_0 with constraints $\{C^*\}$. The resulting mobility matrix for native densities in Eq. (11) is assumed to be the same as that which arises from the Mori–Zwanzig projected dynamics.^{52,53} We begin by outlining this formalism as it was applied in Zwanzig’s classic paper on generalized Rouse dynamics for a polymer described by an arbitrary potential.² A closed form of the mobility matrix is then obtained by approximating the local dynamics with those of the constrained polymer Hamiltonian, H_0 .

Consider a polymer described by the potential $U(\mathbf{R})$ with monomer positions $\mathbf{R}=(\mathbf{r}_1, \dots, \mathbf{r}_n)$. The probability density of monomer positions, $\Psi(\mathbf{R}, t)$ is assumed to evolve according to the Smoluchowski equation

$$\partial_t \Psi(\mathbf{R}, t) = \mathcal{D} \Psi(\mathbf{R}, t), \quad (19)$$

$$\mathcal{D} = \sum_{ij} \nabla_i \cdot D_{ij} e^{-\beta U(\mathbf{R})} \nabla_j e^{\beta U(\mathbf{R})}, \quad (20)$$

where $\nabla_i \equiv \partial/\partial \mathbf{r}_i$, and D is the diffusion matrix. In this section, we denote the equilibrium averages by $\langle \dots \rangle = \int d\mathbf{R} \dots \Psi_{\text{eq}}(\mathbf{R})$ where

$$\Psi_{\text{eq}}(\mathbf{R}) = e^{-\beta U(\mathbf{R})} \left[\int d\mathbf{R} e^{-\beta U(\mathbf{R})} \right]^{-1}. \quad (21)$$

Now consider a set of dynamical variables that are functions of the monomer positions, $\{A_i(\mathbf{R})\}$. These variables will be chosen to be the order parameters $\{\rho\}$ to describe the barrier crossing dynamics discussed in the preceding section, but for now $\{A\}$ is arbitrary. The equilibrium time dependent correlation functions of

$$\delta A_i = A_i - \langle A_i \rangle \quad (22)$$

are denoted by

$$C_{ij}^A(t) \equiv \langle \delta A_i(t) \delta A_j(0) \rangle. \quad (23)$$

The time dependent correlations can be formally expressed as

$$\begin{aligned} C_{ij}^A(t) &= \int d\mathbf{R} \delta A_i e^{\mathcal{L}t} (\delta A_j \Psi_{\text{eq}}) \\ &= \int d\mathbf{R} \Psi_{\text{eq}} \delta A_i e^{\mathcal{L}t} \delta A_j. \end{aligned} \quad (24)$$

In the second line of Eq. (24), we have introduced the adjoint operator \mathcal{L} defined by its action on an arbitrary function B

$$\mathcal{D} \Psi_{\text{eq}} B = \Psi_{\text{eq}} \mathcal{L} B, \quad (25)$$

or explicitly,

$$\mathcal{L} = \sum_{ij} e^{\beta U(\mathbf{R})} \nabla_i D_{ij} e^{-\beta U(\mathbf{R})} \cdot \nabla_j. \quad (26)$$

The meaning of the operator notation in Eq. (24) is defined through the propagator (or Green’s function) $P(\mathbf{X}t; \mathbf{Y})$:

$$C_{ij}^A(t) = \int d\mathbf{X} d\mathbf{Y} \delta A_i(\mathbf{X}) \delta A_j(\mathbf{Y}) P(\mathbf{X}t; \mathbf{Y}) \Psi_{\text{eq}}(\mathbf{Y}), \quad (27)$$

where (formally)

$$P(\mathbf{X}t; \mathbf{Y}) = e^{\mathcal{L}t} \delta(\mathbf{X} - \mathbf{Y}), \quad (28)$$

i.e., $P(\mathbf{X}t; \mathbf{Y})$ satisfies the Smoluchowski equation, $\partial_t P = \mathcal{D}P$, with initial conditions $P(\mathbf{X}t=0; \mathbf{Y}) = \delta(\mathbf{X} - \mathbf{Y})$.

An alternative equation can be derived for the correlations through the projected dynamics of Eq. (19). Following Ref. 2, we define the projection operator through its action on a arbitrary function B

$$\mathcal{P}B = \sum_{ij} \delta A_i \Gamma_{ij}^A \langle \delta A_j B \rangle, \quad (29)$$

where $(\Gamma^A)^{-1}$ is the matrix of static correlations

$$(\Gamma^A)^{-1} = \langle \delta A_i \delta A_j \rangle. \quad (30)$$

Using standard projection operator techniques,^{52,53} it can be shown that the correlations obey the generalized Langevin equation²

$$\partial_t C^A(t) = -\Omega^A \Gamma^A C^A + \int_0^t dt' K^A(t-t') \Gamma^A C^A(t'), \quad (31)$$

where

$$\Omega_{ij}^A = -\langle A_i \mathcal{L} A_j \rangle = \sum_{kl} \langle \nabla_k A_i \cdot D_{kl} \cdot \nabla_l A_j \rangle \quad (32)$$

and $K(t)$ is the memory kernel

$$K_{ij}^A(t) = \langle A_i \mathcal{L} (1 - \mathcal{P}) \exp[(1 - \mathcal{P}) \mathcal{L} t] (1 - \mathcal{P}) \mathcal{L} A_j \rangle. \quad (33)$$

Equation (31) can be written in the form of Eq. (12)

$$\partial_t \mathcal{C}^A(t) = - \int_0^t dt' \mu^A(t-t') \Gamma^A \mathcal{C}^A(t'), \quad (34)$$

where

$$\mu^A(t) = 2\Omega^A \delta(t) - K^A(t) \quad (35)$$

is the effective time-dependent mobility matrix. As suggested by this notation, the static correlations $(\Gamma^A)^{-1}$ can be thought of as arising from an effective harmonic free energy $\beta F = \text{const} + (1/2) \sum_{ij} \delta A_i \Gamma_{ij}^A \delta A_j$.

Equation (34) is an exact result, equivalent to Eq. (27). The linear response formulation avoids explicit use of the Green's function which is generally unknown for an arbitrary potential. On the other hand, it is very difficult to calculate the memory kernel explicitly from Eq. (33) since the projection operators are very awkward to manipulate.

If we choose $\{A\}$ to be the monomer positions, then Eq. (31) becomes an equation for the monomer correlation. If one neglects the memory term, the monomer correlations are said to obey optimized Rouse dynamics.^{2,3} This is equivalent to having an effective preaveraged diffusion matrix, $\Omega = \langle D \rangle$, and a Gaussian approximation to the chain through the static correlations. (Bixon and Zwanzig first presented the stiff chain connectivity matrix [Eq. (4) in paper I] in a paper that employs the optimized Rouse dynamics formalism.) One usually must resort to approximate methods to include the memory kernel in the analysis. A clear presentation of these methods applied to one-dimensional problems is given in Ref. 54.

One way to study the memory function for monomer correlations is through a truncated Mori continued fraction.⁵⁵ This becomes rapidly more cumbersome as the order of the expansion is increased. A more convenient expansion of the memory kernel is to first expand the Green's function in Eq. (27) in the eigenfunctions of \mathcal{L} , and then expand these eigenfunction in a finite basis (similar to what is done in quantum chemistry calculations).⁵⁶⁻⁶² The basis functions must be chosen with some care for accurate results. Furthermore, both the continued fraction and eigenfunction expansion methods require equilibrium averages that are difficult to compute for an arbitrary potential, requiring simulations in general. Nevertheless, this is a useful method since the simulation time needed to compute the equilibrium averages can be much shorter than would be required to simulate the time-dependent correlation functions themselves.⁵⁷

To connect to the barrier crossing calculation for protein folding, we need to study the dynamics of the native density, $\{A\} = \{\rho\}$. For a general potential, calculating the correlation functions $\langle \delta \rho_i(t) \delta \rho_j(0) \rangle$ involves the same technical difficulties that arises for the calculation of the monomer correlations. We avoid these complications by approximating the potential by the constrained polymer Hamiltonian H_0 . Not

only is this a reasonable choice given the interpretation of the order parameters as averages over H_0 , it also allows us to calculate the correlation from Eq. (27) directly.

For an harmonic potential, the propagator in Eq. (27) can be solved analytically.⁶³ The Gaussian form of $\rho(\mathbf{r}_i)$ allows the integrals in Eq. (27) to be performed to give the native density correlations (see Appendix for details)

$$\langle \rho_i(t) \rho_j(0) \rangle_0 = (\det M_{ij}(t))^{-3/2} \times \exp \left[- \frac{3\alpha^N}{2a^2} \mathbf{J}_{ij}^T \cdot M_{ij}(t)^{-1} \mathbf{J}_{ij} \right], \quad (36)$$

where

$$M_{ij}(t) = \begin{bmatrix} (1 + \alpha^N G_{ii}) & \alpha^N G_{ij}(t) \\ \alpha^N G_{ij}(t) & (1 + \alpha^N G_{jj}) \end{bmatrix}, \quad (37)$$

and, \mathbf{J}_{ij} is the two-component difference vector

$$\mathbf{J}_{ij} = \begin{bmatrix} \mathbf{r}_i^N - \mathbf{s}_i \\ \mathbf{r}_j^N - \mathbf{s}_j \end{bmatrix}. \quad (38)$$

In this expression, we have introduced the time correlation function of monomer positions

$$G_{ij}(t) = \langle \delta \mathbf{r}_i(t) \cdot \delta \mathbf{r}_j(0) \rangle_0 / a^2. \quad (39)$$

The chain dynamics are contained in $G_{ij}(t)$, and the precise form depends on the diffusion matrix D_{ij} in Eq. (20). We postpone further discussion of $G_{ij}(t)$ until the next section.

The correlations in Eq. (27) for $\{A\} = \{\rho\}$ are written as

$$\mathcal{C}_{ij}(t) = \langle \delta \rho_i(t) \delta \rho_j(0) \rangle_0 = \langle \rho_i(t) \rho_j(0) \rangle_0 - \rho_i \rho_j, \quad (40)$$

where the first term is given by Eq. (36)–(38) and the second term is found from Eq. (5).

Since approximating the chain potential with H_0 permits an exact solution for the correlation functions, it may seem as if the discussion of the projected dynamics was an unnecessary detour. However, what we are really trying to calculate is the effective mobility matrix needed to evaluate the folding rate prefactor in Eq. (16).

From Eq. (34), the native density correlations obey the projected relaxation equation

$$\partial_t \mathcal{C}(t) = - \int_0^t dt' \mu^\rho(t-t') \Gamma^\rho \mathcal{C}(t'), \quad (41)$$

where Γ^ρ are the inverse of the static correlations

$$[\Gamma^\rho]_{ij}^{-1} = \langle \delta \rho_i \delta \rho_j \rangle_0 = \mathcal{C}_{ij}(0). \quad (42)$$

Equation (35) with $\{A\} = \{\rho\}$ gives $\mu^\rho(t)$, in terms of projection operators, but we can also deduce an alternative expression for the mobility since $\mathcal{C}(t)$ is known. Solving Eq. (41) for $\mu^\rho(t)$ by Laplace transforms gives

$$\hat{\mu}^\rho(\omega) = \mathcal{C}(0) \hat{\mathcal{C}}(\omega)^{-1} \mathcal{C}(0) - \omega \mathcal{C}(0). \quad (43)$$

To close this expression for the mobility, we only need the Laplace transform of $\mathcal{C}(t)$ which can be calculated numerically.

As discussed previously, Eq. (41) is consistent with a Gaussian approximation to the free energy

$$\beta F^p = \text{const} + \frac{1}{2} \sum_{ij} \Gamma_{ij}^p \delta \rho_i \delta \rho_j \quad (44)$$

analogous to the saddle-point expansion of $\bar{F}[\{\rho^*\}]$ given in Eq. (8). However, even though Γ^p can be evaluated at the saddle-point constraints in $\{C^*\}$ in H_0, F^p is stable against all fluctuations away from $\delta \rho_i = 0$. This is in contrast to the curvature matrix $\bar{\Gamma}^*$ that has one unstable direction.

Nevertheless, by approximating the mobility in Eq. (12) by $\mu(t) \approx \mu^p(t)$ we can solve the eigenvalue equation (16) to identify the unstable mode \mathbf{u}^* and calculate the prefactor κ^* . The last remaining unspecified quantity for the theory is the time dependent monomer correlations $G(t)$ which is the subject of the next section.

IV. CHAIN DYNAMICS

In this section we specify our model for the chain dynamics, and calculate the monomer pair correlation, $G(t)$, with the approximation to the chain potential $U(\mathbf{R}) = H_0$. This Gaussian approximation simplifies the true microscopic potential that governs conformational changes of the protein chain, and by itself ignores explicit interactions. We attempt to capture some effects of the microscopic dynamics by including frictional forces that are independent of the solvent viscosity, i.e., internal viscosity.

A. Monomer pair correlations

Starting from the definition of pair correlations for the Smoluchowski equation [Eq. (27)], it is easy to derive an equation for the relaxation of the monomer correlations

$$\partial_t \langle \delta \mathbf{r}_i(t) \cdot \delta \mathbf{r}_j(0) \rangle = \sum_k \gamma_{ik}^{-1} \langle \nabla_k U(\{\mathbf{r}(t)\}) \cdot \delta \mathbf{r}_j(0) \rangle, \quad (45)$$

where the friction coefficient is related to the diffusion coefficient by

$$\beta D = \gamma^{-1}. \quad (46)$$

In Eq. (45), we have assumed that γ is independent of $\{\mathbf{r}\}$, but may still depend on the sequence index. The model for the friction matrix is specified below, for now we define a dimensionless friction matrix $\tilde{\gamma}$:

$$\gamma = \gamma_0 \tilde{\gamma}, \quad \gamma_0 = 6 \pi a_{\text{eff}} \eta, \quad (47)$$

where we have used Stokes law to define γ_0 with the solvent viscosity η and typical monomer van der Waals radius a_{eff} .

For a given point along the folding route (e.g., a transition state), we assume that the chain dynamics can be described by the reference Hamiltonian [Eq. (13) in paper I]. Physically, H_0 corresponds to a polymer confined to the native structure by harmonic constraints with spring constants $\{3/2a^2 C_i\}$. These constraints act as a nonuniform external field, controlling the fluctuations about the native structure. Taking the gradient of H_0 , and using the definition [Eq. (15) in paper I] of the average monomer position, \mathbf{s}_i , gives

$$\begin{aligned} \frac{\partial H_0}{\partial \mathbf{r}_i} &= \frac{3k_B T}{a^2} \left[\sum_j \Gamma_{ij}^{(\text{ch})} \mathbf{r}_j + C_i (\mathbf{r}_i - \mathbf{r}_i^N) \right] \\ &= \frac{3k_B T}{a^2} \sum_j \Gamma_{ij}^{(0)} \delta \mathbf{r}_j, \end{aligned} \quad (48)$$

where $\Gamma_{ij}^{(0)}$ is given in Eq. (17) in paper I and $\delta \mathbf{r}_i = \mathbf{r}_i - \mathbf{s}_i$ as usual. Substituting this into Eq. (45) gives the matrix equation for the correlations $G_{ij}(t) = \langle \delta \mathbf{r}_i(t) \cdot \delta \mathbf{r}_j(0) \rangle / a^2$,

$$\partial_t G(t) = \sigma \tilde{\gamma}^{-1} \Gamma^{(0)} G(t), \quad (49)$$

where

$$\sigma = \frac{3k_B T}{a^2 \gamma_0} = \frac{3D_0}{a^2}. \quad (50)$$

sets the time scale for the relaxation.

Since Eq. (49) is linear, it can be solved by transforming to normal modes by a similarity transform

$$Q^{-1} [\tilde{\gamma}^{-1} \Gamma^{(0)}] Q = \text{diag}\{\lambda_p\}, \quad (51)$$

where the columns of Q are the right eigenvectors of $[\tilde{\gamma}^{-1} \Gamma^{(0)}]$. Even though $\tilde{\gamma}$ and $\Gamma^{(0)}$ are symmetric, their product is not symmetric in general. Nevertheless, Q can diagonalize $\tilde{\gamma}$ and $\Gamma^{(0)}$ separately (though not by a similarity transformation since Q is not orthogonal in general).⁶⁴ With

$$Q^T \tilde{\gamma} Q = \text{diag}\{\nu_p\}, \quad Q^T \Gamma^{(0)} Q = \text{diag}\{\eta_p\} \quad (52)$$

we have the relationship

$$\lambda_p = \eta_p / \nu_p. \quad (53)$$

The correlations between monomers, expanded onto the normal modes is

$$G_{ij}(t) = \sum_p Q_{pi} Q_{jp}^T / \eta_p \exp(-\sigma \lambda_p t). \quad (54)$$

We note that equal time correlation functions are the equilibrium correlations $G_{ij}(t=0) = G_{ij} \equiv [\Gamma^{(0)}]^{-1}$.

The reference Hamiltonian has two features that distinguish this chain model from the nearest neighbor Rouse chain: chain stiffness, parametrized by g , and constraints to the native structure, specified by $\{C\}$. We consider the effect on the pair correlations of these two contributions separately, to give a point of reference for the model. To simplify this illustration of the dynamics, we consider a (free-draining) diagonal friction matrix, $\tilde{\gamma}_{ij} = \delta_{ij}$. Then, $\nu_p = 1$ which implies that the relaxation rates are $\lambda_p = \eta_p$ and the normal modes are the eigenvectors of the chain connectivity $\Gamma^{(\text{ch})}$.

For an unconstrained polymer ($C_i = 0$), the dynamics of the stiff, freely rotating chain has been studied quite extensively.^{3,29-32} Referring to the definition of Γ [Eq. (4) in paper I], Q differs from that for the well-known Rouse modes Q^R ,

$$[Q^R]^T K^R Q^R = \text{diag}\{\eta_p^R\}, \quad (55)$$

because of the boundary term, $g^2/(1-g)\Delta$, which is only important for very stiff chains.³⁰ Thus, unless $g \approx 1$, $Q \approx Q^R$ for the free-draining, unconstrained polymer:

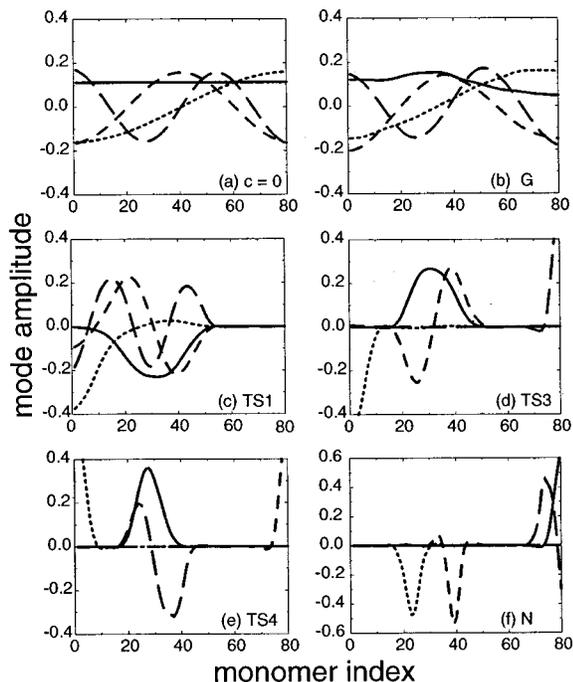


FIG. 1. Mode amplitude plotted vs sequence index for the four modes with the slowest relaxation rates; from the slowest mode: solid, dashed, long dashed, dotted. ($\gamma_I=0$). The constraints correspond to stationary points on the free energy surface for λ -repressor with persistence length $l=5a$ ($g=0.8$): (a) unconstrained, (b) globule, (c) TS_1 , (d) TS_3 , (e) TS_4 , (f) native.

$$Q_{pi}^R \sim \cos\left[\frac{\pi p}{n}(i-1/2)\right], \quad \eta_p^R = 4 \sin^2\left[\frac{\pi p}{2n}\right], \quad (56)$$

where the mode index takes values $p=(0, \dots, n-1)$. The relaxation spectrum is then approximately

$$\lambda_p = \eta_p \approx \frac{1-g}{1+g} \eta_p^R + \frac{g}{1-g^2} (\eta_p^R)^2 + B, \quad (57)$$

where the final term is due to the confinement term in Eq. (9) in paper I. The second term of Eq. (57) causes the short wavelength modes (high mode index p) to relax faster as the chain stiffness increases. In contrast, the relaxation rates for the long wavelength modes (low mode index p) rates are suppressed by increasing chain stiffness. The number of modes with slowly relaxing rates that are suppressed diminishes as the chain stiffness increases; for the limiting stiffness corresponding to a rigid rod, $g \rightarrow 1$, the relaxation rates for all modes except one diverge.³

When the polymer is constrained by a nonuniform pattern of nonzero constraints $\{C_i\}$, the normal modes are not the Rouse modes of a free chain. Figure 1 shows the four slowest modes for a chain with stiffness parameter $g=0.8$ (persistence length $l=5a$) and various constraints. In addition to the modes for the unconstrained chain shown in Fig. 1(a), the modes corresponding to selected stationary points along the folding route of λ -repressor are also shown in Figs. 1(b)–1(f). The slowest modes for the denatured state [Fig. 1(b)] differ only slightly from those for the unconstrained chain [Fig. 1(a)] given by Eq. (56). Referring to the fluctuations shown in Fig. 9(b) of paper I, the structure at the transition state TS_1 can be roughly described by saying that resi-

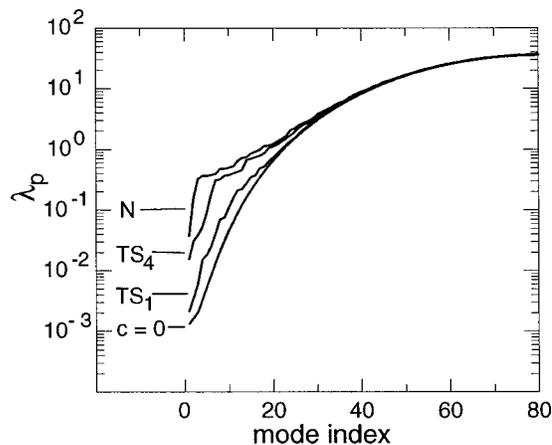


FIG. 2. Relaxation rates λ_p vs mode index for different constraints corresponding to stationary points on the free energy surface for λ -repressor with persistence length $l=5a$ ($g=0.8$): unconstrained ($C=0$), TS_1 , TS_4 , and native (N) are indicated in the plot.

dues in helices H4–H5 (indices 53–80) are localized, while the residues in helices H2–H3 (indices 27–47) have the largest fluctuations.⁶⁵ These constraints are reflected in the structure of the slowest relaxation modes shown in Fig. 1(c). Participation of the more localized residues is shifted to modes with higher relaxation rates so that these residues are absent in the slowest relaxation modes. In addition, the mode with the slowest relaxation (solid line) is no longer a uniform translation mode, but involves motion of the most unconstrained residues in helices H2–H3 (indices 27–47). The other modes are much the same as those of an unconstrained but shorter chain with one end fixed at the end of helix H4. Similar correspondence between the mode structure and the localization of residues in the transition state ensemble can be seen for the constraints corresponding to TS_3 and TS_4 in Figs. 1(d) and 1(e) where the modes become increasingly more localized in sequence due to nonsequential constrained residues. The normal modes for the native state [Fig. 1(f)] are more localized with the slowest relaxation involving localized regions with the largest temperature factors.

The mode spectra (relaxation rates) for the free chain, constraints corresponding to transition states, and the native state are shown in Fig. 2. As can be seen in this figure, while the spectrum for the short wavelength modes (large mode index) are relatively unaffected, the relaxation rates for all the long wavelength modes (small mode index) become faster as the protein takes on more structure, reflecting the relatively rapid motion toward the native structure as the constraints provided by the native structure become stronger.

B. Internal friction

One can expect that Eq. (49) with harmonic forces can capture the correct physics of the low frequency spectrum of the chain dynamics, but, as in the Debye theory of crystals, the description will not be correct for small length scales which depend on microscopic details. For many polymer physics applications, even though the polymer has a broad range of relaxation rates, the experimentally observed dynamics is dominated by the long wavelengths (in sequence)

motion corresponding to the slowest modes. de Gennes calls the independence of the long wavelength modes on the microscopic conformational changes the ‘‘Kuhn theorem.’’⁴² Why then is it interesting to incorporate internal viscosity (arising from local activated isomerization) into a model of barrier crossing dynamics? Because we do not know *a priori* the scale of the motions involved in locally crossing the transition region.

To be more specific, we consider internal friction for the following two reasons: (1) We have seen that the nonuniform constraints to the native structure affects the structure of the normal modes moving them to involve more local motions. It is unclear then how the Kuhn theorem (which is a statement about the delocalized continuum limit) applies here. (2) We are ultimately interested in studying the barrier crossing rate for protein folding, through the Laplace transform of the native density correlations [Eq. (36)] which mix different Rouse modes. A correlation function similar to the native density correlation function arises (for an unconstrained polymer model) in the Wilemski–Fixman theory of diffusion-limited chain closure reactions.^{66,67} In analyzing this theory, Doi showed that for end-to-end monomers, the Laplace transform evaluated at small frequencies is actually dominated by the integral at short times, i.e., all modes contribute to the integral, not just those with the longest relaxation times.⁶⁸

The friction matrix in Eq. (49) $\gamma = \gamma_0 \tilde{\gamma}$ is written as a sum of two contributions

$$\gamma = \gamma^S + \gamma^{IV}, \quad (58)$$

where the first term is due to the solvent viscosity, and the second term accounts for the ‘‘internal viscosity’’ and is taken to be independent of the solvent friction. A typical choice for the effective solvent friction is the Oseen tensor from the hydrodynamic mobility.¹ This nonlocal friction mediated by the solvent damps the motion of one bead due to the motion of another distant bead. It is not difficult to use standard preaveraging approximations to incorporate these hydrodynamic interactions. For simplicity, we neglect this aspect of the problem and consider the free-draining approximation

$$\gamma_{ij}^S = \gamma_0 \delta_{ij}, \quad (59)$$

where γ_0 is defined in Eq. (47).

As previously discussed, some internal friction is due to the neglect of the activated character of the individual conformational transitions of the backbone angles. While there are other ways to incorporate this time scale into a model of chain dynamics,⁶⁹ perhaps the most simple correction to the Gaussian description is through the friction matrix. In accordance with the Kuhn theorem, the long wavelength motion should have the usual solvent viscosity dependence being largely unaffected by the microscopic barrier crossing; on the other hand, short wavelengths should tend to relax more slowly, due to the ‘‘dynamical stiffness’’ caused by the slow activation rates for dihedral angle changes. Since the dihedral angle potentials are quite steep these will have a much weaker viscosity dependence.

In the model presented in Ref. 70, Bazúa and Williams derived an effective frictional force on the *i*th bead which acts on the relative velocities of adjacent beads:

$$F_{IV} = \gamma_I (\dot{\mathbf{r}}_{i+1} - \dot{\mathbf{r}}_i)^d - \gamma_I (\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_{i-1})^d, \quad (60)$$

where the superscript denotes the projection of the forces along the bond directions. A common approximation is to write the projected velocities as a sum of the difference between the full velocity vectors $\{\dot{\mathbf{r}}\}$ and the difference between angular velocity of the beads.⁷¹ The angular component is important for inherently anisotropic phenomena such as flow birefringence, but for our purposes we can neglect the angular rotation component and write the internal friction in the form^{39–41,70}

$$F_{IV} = -\gamma_I \sum K_{ij}^R \partial_i (\delta \mathbf{r}_j), \quad (61)$$

where K^R denotes the Rouse connectivity matrix [Eq. (5) in paper I], and we have used $\partial_i (\delta \mathbf{r}_j) = \partial_i \mathbf{r}_j$ since the average monomer position is constant for fixed constraints.

The total friction matrix is the sum of the (free-draining) diagonal solvent friction and the internal friction

$$\begin{aligned} \gamma_{ij} &= \gamma_0 \delta_{ij} + \gamma_I K_{ij}^R \\ &= \gamma_0 (\delta_{ij} + \gamma_I / \gamma_0 K_{ij}^R), \end{aligned} \quad (62)$$

which defines the dimensionless friction matrix $\tilde{\gamma}$ in Eq. (47).

The spectrum of relaxation rates with nonzero internal friction can be illustrated by considering an unconstrained polymer chain ($C_i = 0$). From Eq. (53) the mode relaxation rates for this case is approximately

$$\lambda_p \approx \frac{\eta_p}{1 + \gamma_I / \gamma_0 \eta_p^R}, \quad (63)$$

where η are the eigenvalues of $\Gamma^{(0)}$ [Eq. (57)] and η_p^R are the eigenvalues of the Rouse matrix K^R . This equation shows that the internal friction leaves the rates of the long wavelength modes (small η^R) relatively unchanged; this is the Kuhn theorem. The short wavelength modes (larger η^R), are suppressed by a nonvanishing internal friction. The crossover between the two behaviors, hydrodynamic solvent limited versus conformational interconversion limited, is determined by the value of γ_I / γ_0 .

For constrained chains, the effect of internal friction on the relaxation modes is different. The internal friction damps rapidly oscillating features of the modes which even occur in the relaxation modes with long relaxation times since they are not identical to Rouse modes. (Note the modes illustrated in Fig. 1, which are also different from Rouse mode were calculated for $\gamma_I = 0$.) Nevertheless, for the constraints considered here, this seems to be a minor effect, and the relaxation modes still follow the general behavior of Eq. (63). Figure 3 shows the relaxation spectrum with zero and finite internal friction for the unconstrained chain and for the constraints corresponding to the transition state TS₂ of the variational calculation for λ -repressor. The spectrum corresponding to the other transition states is similar: the relaxation

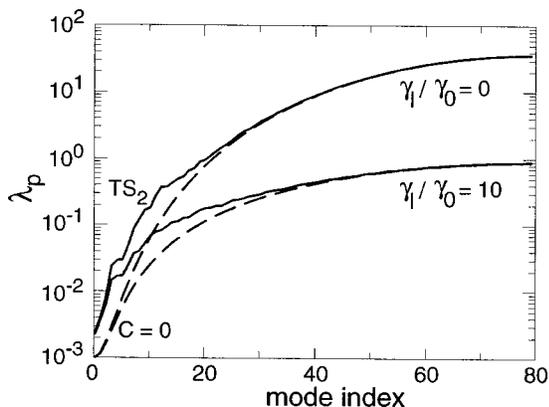


FIG. 3. Relaxation rates λ_p vs mode index for two values of internal viscosity: $\gamma_1/\gamma_0=0$ and $\gamma_1/\gamma_0=10$ (indicated on the plot). The constraints correspond to λ -repressor with persistence length $l=5a$ ($g=0.8$): unconstrained (dashed lines) and TS_2 (solid lines).

rates for the slow modes are relatively unaffected, whereas the faster modes have suppressed relaxation rates.

V. FOLDING PREFACTOR EXAMPLE: λ -REPRESSOR

The average folding route for the fast folding variant of the λ -repressor protein presented in paper I is characterized by sets of constraints $\{C\}$ corresponding to local minima and saddle points of the variational free energy. For each saddle point, the pair correlations $G(t)$ are calculated through the normal mode expansion given by Eq. (54) with the friction matrix given by Eq. (62). These monomer correlations determine the native density correlations $\mathcal{C}(t)$ through Eq. (40) [with Eq. (5) and Eq. (36)], and its Laplace transform $\hat{\mathcal{C}}(\omega)$ is taken numerically. With the mobility matrix $\hat{\mu}(\omega)$ given by Eq. (43) and the curvature matrix of the free energy $\bar{\Gamma}^*$ given by Eq. (10), the eigenvalue equation (16) is then solved numerically giving the reaction coordinate and folding rate prefactor for that transition state.

In this example, we consider the folding routes obtained for a chain with persistence length of polyalanine, $l=5a$ ($g=0.8$). The native density $\{\rho\}$ also depends on the parameter α^N controlling the width of the Gaussian measure to the native structure. This width should be small enough to characterize the native structure accurately, but the appropriate value is also limited by the resolution of the model [approximately equal to the root mean square (RMS) fluctuations of the bond length a]. For large α^N (small width), $\{\rho\}$ has low values even for the native state since the average positions are not precisely equal to the native structure at finite temperature; for $\alpha^N=1$ the average value of ρ_i evaluated at the native state is 0.3. For the native densities shown in Fig. 4 with $\alpha^N=0.5$, the average of ρ_i at the native state increases to ρ_i is 0.5. We consider this to be a reasonable value to use for α^N . For smaller α^N (larger widths), the separation between $\{\rho\}$ evaluated at the native and globule state is more distinct, though unfolded regions of the transition states have native densities closer to the native values. The dependence of the prefactor on the width α^N is considered below.

The local reaction coordinates (unstable modes) for the four transition states along the folding route are shown in

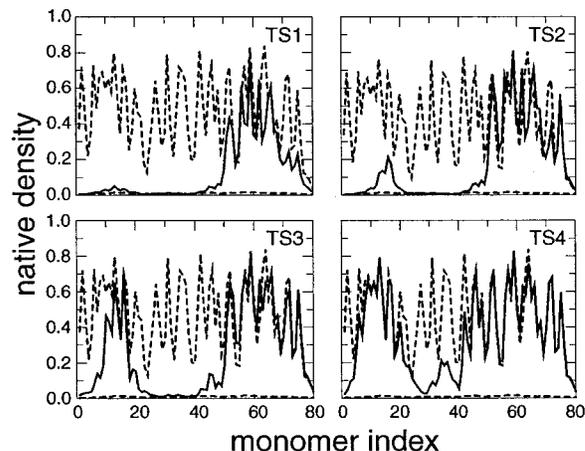


FIG. 4. Native density ρ_i evaluated at the native and globule minima (dashed) and the transition states TS_1 – TS_4 (solid) of the average folding route for λ -repressor. The persistence length of the chain is $l=5a$ ($g=0.8$) and the width defining ρ_i is set by $\alpha^N=0.5$.

Fig. 5. These unstable modes are quite understandable in terms of the structure inferred by the temperature factors in Fig. 9(b) of paper I. The unstable mode at TS_1 is rather delocalized, consisting of residues in helices H4–H5 and modulates with a wavelength approximately equal to the persistence length $l/a=5$; the partial localization of helix H1 indicated by the temperature factors is not a part of the normal mode. The unstable mode at TS_2 has a significant component of formation of helices H4–H5 as in TS_1 , but the localized region around a residue in H1 is a much larger component. The unstable mode at TS_3 is also localized, but consists mainly of a few residues separated in sequence in helix H1 and near helix H4. Finally, the unstable mode at TS_4 is delocalized consisting of residues in helices H2–H3 which are the last helices to form. The features of the reaction coordinates resemble the unstable modes of the free energy ($\bar{\Gamma}$), though the relative amplitudes of the peaks in a given reaction coordinate are altered by the chain dynamics through the mobility $\hat{\mu}(\omega)$.

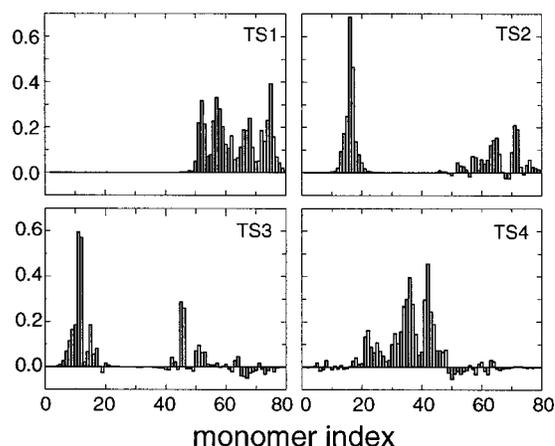


FIG. 5. Local reaction coordinates for the transition states TS_1 – TS_4 of the λ -repressor average folding route at T_f . Normalization of the reaction coordinate is set to $\sum \mathbf{u}_i^2=1$. Parameters are the same as in Fig. 4.

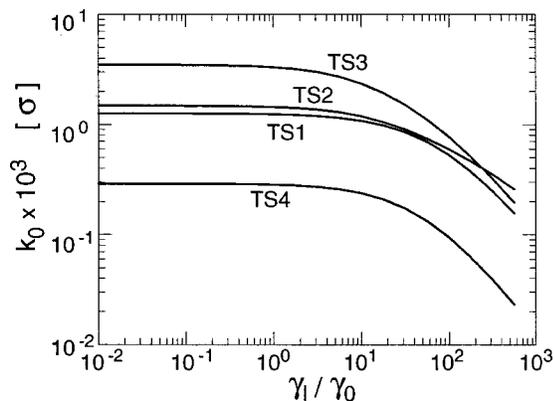


FIG. 6. Prefactor $k_0 = |\kappa^*|/2\pi$ in units of $\sigma = 3D_0/a^2$ as a function of γ_1/γ_0 for the transition states TS₁–TS₄ of the λ -repressor average folding route at T_f . Parameters are the same as in Fig. 4.

Since the local reaction coordinates along the average folding route are very different, one from each other, a reduced description of the kinetics based on a single reaction coordinate cannot be used when dynamics is examined at this level of detail. Attempts to find perfect global dynamical coordinates in lattice simulations of perfectly funneled models are consistent with this.^{22,72} In the present description, each dynamical reaction coordinate applies in the vicinity of one saddle point. The changes of the local reaction coordinate along the folding route are intimately related to the fine structure of the free energy profile. Notice that the effects on predicted reaction rates are minor, however. Surmounting the highest barrier gives the rate within a factor of 2 (*vide supra*). Calculations of one-dimensional barrier crossing with shallow local minima (intermediates) also show why having a perfectly good local detailed description of the free energy profile need not yield a strictly defined global reaction coordinate.⁷³ At the same time, a much cruder reaction coordinate can function perfectly well for predicting rates if the appropriate configurational diffusion coefficient is used.⁷⁴

The prefactor $k_0 = |\kappa^*|/2\pi$ for each transition state is plotted as a function of the internal friction γ_1/γ_0 in Fig. 6. When the unstable curvature of the free energy is large, there is less recrossing at the top of the barrier leading to a larger prefactor. For small γ_1/γ_0 the prefactors increase in the same order as the corresponding curvatures of the free energy ($\bar{\Gamma}_0$): $k_0^{(4)} < k_0^{(1)} < k_0^{(2)} < k_0^{(3)}$. The relative values of the prefactors at different transition states depend on the chain dynamics through the mobility matrix $\hat{\mu}(\omega)$, however, and do not follow the proportions of the corresponding free energy curvatures. For small values of γ_1/γ_0 , the prefactor is proportional to the bare diffusion coefficient $D_0 = k_B T/\gamma_0$. The prefactors decrease with increasing γ_1/γ_0 because of the suppressed relaxation rates of the fastest modes. This turnover occurs around $\gamma_1/\gamma_0 = O(1)$, though it is different for each transition state.

To put these results in laboratory units for polypeptides, we need to specify the monomer diffusion coefficient D_0 and the internal viscosity parameter γ_1/γ_0 . We estimate the typical hydrodynamic radius in Eq. (47) to be the sum of the monomer spacing $a = 3.8 \text{ \AA}$ and the average van der Waals

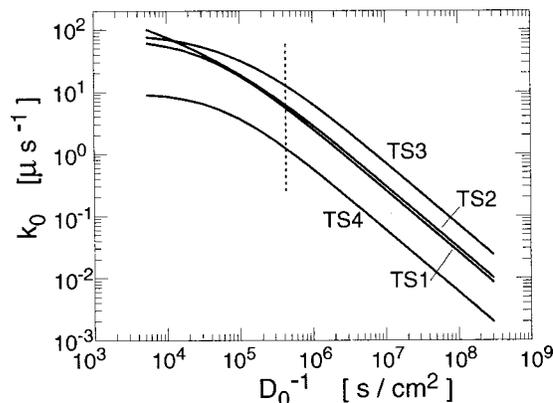


FIG. 7. Prefactor k_0 is plotted as a function of the inverse monomer diffusion coefficient $D_0^{-1} = \gamma_0/k_B T_f$. The internal friction is set equal to the fixed value $\gamma_1/k_B T_f = 9.1/D_0^{\text{H}_2\text{O}}$. The vertical dotted line corresponds $D_0 = D_0^{\text{H}_2\text{O}} = 3 \times 10^{-6} \text{ cm}^2/\text{s}$ (and $\gamma_1/\gamma_0 = 9.1$). Parameters are the same as in Fig. 4.

radius of a typical side chain⁷⁵ $a_{\text{vdw}} \approx 3 \text{ \AA}$; $a_{\text{eff}} \approx 7 \text{ \AA}$. Using the viscosity of water at room temperature ($\eta_0 = 1$ centipoise) gives the a monomer diffusion coefficient $D_0^{\text{H}_2\text{O}} = 3 \times 10^{-6} \text{ cm}^2/\text{s}$. We estimate the internal viscosity by the elementary bond flipping rate of the polymer backbone. Calculations of the $\beta \rightarrow \alpha$ conformations of di-alanine give a value for $\tau_{\text{flip}}^{-1} \approx 10 \text{ ns}^{-1}$.⁷⁶ One can expect this rate to be reduced when the flipping transition occurs in a long polypeptide chain since the neighboring monomers must adjust to some extent in order to minimize the motion of the chain through the solvent.²⁷ Judging from studies on model chains one expects about a factor of 3 slower,^{27,28} we estimate a typical value to be $\tau_{\text{flip}}^{-1} = 3 \text{ ns}^{-1}$. It would be nice if simulations could pin this number down for the peptide isomerization rate in long chains. It is obvious that the relaxation rates describing the chain dynamics should not exceed this elementary rate. Associating the internal friction with this isomerization rate, we set γ_1/γ_0 so that the free chain relaxation rates of the fastest mode equals τ_{flip}^{-1} . For a chain with persistence length $l = 5a$ and monomer diffusion coefficient $D_0^{\text{H}_2\text{O}}$, we find $\gamma_1/\gamma_0 = 9.1$.

In Fig. 7, the prefactor k_0 is plotted against the inverse monomer diffusion coefficient D_0^{-1} in laboratory units. The internal friction is set to the fixed value $\gamma_1 = 9.1\gamma_0$ for $\gamma_0/k_B T_f = (D_0^{\text{H}_2\text{O}})^{-1}$. The prefactor is proportional to the solvent viscosity ($D_0^{-1} \sim \eta_0$) for large viscosities and saturates at low viscosities. This turnover is reminiscent of the viscosity dependence of the prefactor from Kramer's theory,⁴⁴ but here the turnover is driven indirectly by the internal friction of the chain rather than explicit inertial terms reflecting an energy limited regime. The prefactor in pure water, indicated by the dashed line in Fig. 7, is still in the linear regime, but near the turnover. The prefactors of the different transition states for $D_0 = D_0^{\text{H}_2\text{O}}$ (given in Table I) are on the order of $10 \mu\text{s}^{-1}$. As a point of reference, the unstable mode frequency for TS₂, $|\kappa^*| = 7.5 \times 10^{-3} \sigma$, is approximately equal to the relaxation rate of the $p = 5$ normal mode of a free chain of $n = 80$ monomers and persistence length $l = 5a$ ($\lambda_{p=5} = 7.2 \times 10^{-3}$). The normal mode for this mo-

TABLE I. Prefactor $k_0 = |\kappa^*|/2\pi$ in units of $\sigma = 3D_0/a^2$ and in μs for λ -repressor at the folding transition temperature, T_f . The parameters are: $l = 5a(g = 0.8)$, $\alpha = 0.5$, $D_0 = 3 \times 10^{-6} \text{ cm}^2/\text{s}$, $\gamma_l/\gamma_0 = 9.1$.

	TS ₁	TS ₂	TS ₃	TS ₄
$k_0 \times 10^3 [\sigma]$	1.1	1.2	2.4	0.24
$k_0 [\mu\text{s}^{-1}]$	6.8	7.6	15	1.5

tion has a node to node period of 16 monomers, roughly twice the length used in the estimates based on the smallest loop closure.^{24,25}

The folding rate includes the activation barrier (calculated with this model in paper I) as well as the prefactor. The calculated thermodynamic and kinetic quantities needed for the folding rate are given in Table II for two thermodynamic conditions. Adjusting the energy scale of the contacts $\epsilon_0/k_B T$ fixes the equilibrium constant $K = k_f/k_u$ (the ratio of the folding and unfolding rates), so that $\ln K$ gives the difference in free energy between the globule and native state (in units of $k_B T$). For the very stabilizing condition $\ln K = 7.8$, the prefactor is $k_0 = 2.4 \mu\text{s}^{-1}$ and the barrier is $\Delta F^\ddagger/k_B T = 3.4$, giving a folding time of $1/k = 13 \mu\text{s}$. This agrees well with the experimental rate $1/k = 12 \mu\text{s}$ at this stability (extrapolated to 0 M denaturant).⁷⁷ At the transition midpoint $\ln K = 0$, the calculated prefactor is $7.6 \mu\text{s}$, and with the barrier height $\Delta F^\ddagger/k_B T = 5.1$, we get a folding time of $1/k = 23 \mu\text{s}$ which is considerably faster than the measured folding time of $1/k \approx 10 \text{ ms}$.⁷⁷ This suggests that the microscopic calculation for the perfectly funneled surface underestimates the barrier by about $5k_B T$ at the folding transition temperature.

The calculated inverse prefactor for λ -repressor at the transition midpoint $1/k_0 \approx 0.1 \mu\text{s}$ and under highly stabilizing condition $1/k_0 \approx 0.5 \mu\text{s}$ are both faster than the estimated fastest possible folding time^{24,25} $\tau \approx 1 \mu\text{s}$. The latter is the time it takes to form a typical contact as determined by the fluorescence or chemical quenching rate of a pair of residues separated by a typical loop length. It is very difficult to imagine that folding can occur on times faster than contact formation, and the introduction of this speed limit is important conceptually as a reference time scale for protein folding dynamics. It is tempting to identify the speed limit with the folding prefactor since the fastest rate is obtained from Eq. (1) for a vanishing free energy barrier. But this is not precisely correct since the separation of time scales assumed in activated kinetics is not true for small barriers. As seen in Table II, reducing the activation barrier (by stabilizing the native state, for example) changes the prefactor since the

TABLE II. Calculated kinetic parameters at two stabilities $\ln K = -\beta(F_N - F_G)$ where F_N and F_G are the free energy of the native and globule states: contact energy $\epsilon_0/k_B T$, barrier height $\Delta F^\ddagger/k_B T$, corresponding folding prefactor $k_0 = |\kappa^*|/2\pi$, and folding time $\tau = k^{-1}$. Other parameters are fixed: $l = 5a(g = 0.8)$, $\alpha = 0.5$, $D_0 = 3 \times 10^{-6} \text{ cm}^2/\text{s}$, $\gamma_l/\gamma_0 = 9.1$.

$\ln K$	$k_B T/\epsilon_0$	$\Delta F^\ddagger/k_B T$	$k_0 [\mu\text{s}^{-1}]$	$\tau [\mu\text{s}]$
0.0	1.44	5.1	7.6	23
7.8	1.30	3.4	2.3	13

TABLE III. Internal friction and prefactor (from TS₂) for λ -repressor for different values of the monomer diffusion coefficient in water $D_0^{\text{H}_2\text{O}}$ and isomerization rate τ_{flip}^{-1} . $D_0^{\text{H}_2\text{O}} = 9.1 \times 10^{-6} \text{ cm}^2/\text{s}$ is the diffusion coefficient of free alanine (Ref. 75) and $\tau_{\text{flip}}^{-1} = 10 \text{ ns}^{-1}$ is the calculated isomerization rate for di-alanine (Ref. 76). The parameters in the first line are considered the best estimates. The persistence length is $l = 5a(g = 0.8)$, the width of the Gaussian measure of the native density is $\alpha^N = 0.5$, and the temperature is the folding transition temperature.

$D_0^{\text{H}_2\text{O}} \times 10^6 (\text{cm}^2/\text{s})$	$\tau_{\text{flip}}^{-1} (\text{ns}^{-1})$	γ_l/γ_0	$k_0 (\mu\text{s}^{-1})$
3	3	9.1	7.6
3	10	2.5	8.7
9.1	3	28	18
9.1	10	8.2	23

saddle-point curvature of the free energy changes. If the prefactor at both stabilities in this example were $1/k_0 = 0.1 \mu\text{s}$, the folding time with the reduced activation barrier would be a factor of approximately 6 times smaller than at the transition midpoint, or $1/k \approx 4 \mu\text{s}$, which approaches the speed limit. The prefactor is found to be approximately three times smaller than at the midpoint, however, resulting in a larger folding time $1/k \approx 12 \mu\text{s}$. Pushing the conditions to the limit of a vanishing barrier, it could happen that the barrier is very low but still has a large unstable mode free energy curvature, resulting in a calculated folding time somewhat faster than the speed limit. For such near downhill folding,^{78,79} however, the folding would not be a single exponential process dominated by barrier crossing, but involves kinetics more directly connected to the time for contact formation than the time reflected in the apparent prefactor.

It is interesting to consider the dependence of the prefactor on the chain dynamical parameters since the values for the monomer diffusion coefficient and the isomerization rate in a polymer chain are not known precisely (see next section). Conservatively, the values we used are probably accurate to within a factor of 2 or 3. As long as the internal friction is not too large, the monomer diffusion coefficient is the more important parameter for the absolute prefactor (in s^{-1}); since the time scale for the monomer relaxation is $3D_0/a^2$, the absolute prefactor is proportional to this scaling. Table III gives the prefactor from TS₂ using the free alanine diffusion coefficient and the di-alanine isomerization rate. As shown in Table III, the prefactor varies by about a factor of 3 when $D_0^{\text{H}_2\text{O}}$ is the diffusion coefficient of a free alanine molecule. This is the same factor as the ratio of the monomer diffusion coefficient, though it also depends on the value of the internal friction (i.e., τ_{flip}^{-1}) as well.

The value of the Gaussian width α^N defining the native density is a parameter of the model, not a measurable quantity, so we must also consider the sensitivity of our results to the value of α^N . As shown in Fig. 8, the prefactors k_0 are approximately linear in α^N , increasing by a factor of about 2 for TS₃ and less than a factor of unity for the other transition states as α^N varies from 0.25 to 1.0. The different slopes reflect the different increases in curvatures of the free energies as the Gaussian width narrows. This degree of uncertainty seems reasonable, but the different slopes suggests that

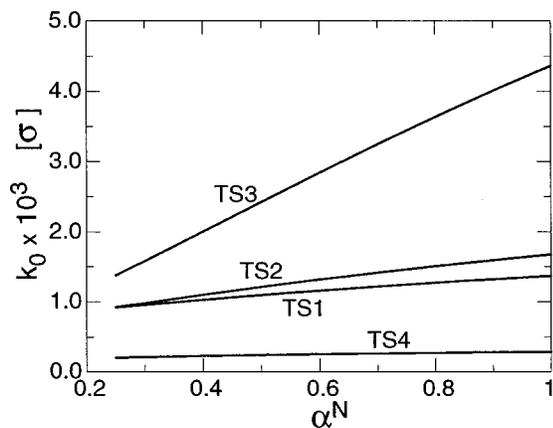


FIG. 8. Prefactor $k_0 = |\kappa^*|/2\pi$ vs the Gaussian width of the native density α^N . The persistence length of the chain is $l = 5a$ ($g = 0.8$) and the internal friction is $\gamma_l/\gamma_0 = 0.91$.

the dependence is a delicate issue, depending on the detailed structure of each transition state.

VI. CONCLUDING REMARKS

In this paper, we derived a microscopic theory for the dynamics of protein folding barrier crossing when non-native contacts and trapping effects can be neglected. The variational theory presented in paper I characterizes the average folding routes as connected minima and transition states. These routes are specified through variational parameters constraining the protein chain to inhomogeneous order about the native structure. The quadratic reference Hamiltonian from this theory along with forces from solvent and internal friction govern the chain dynamics near the saddle points. The barrier crossing dynamics is calculated from a generalization of both Grote–Hynes and Langer rate theories to include a frequency dependent friction in multidimensions. The memory friction itself is approximated by linear response theory and exploits the harmonic potential of the model. The theory gives local reaction coordinates and folding rate prefactors for specific proteins with known native structure.

For the calculated average folding route of λ -repressor, we find a folding prefactor $k_0 \approx 10 \mu\text{s}^{-1}$ at the transition midpoint and $k_0 \approx 1 \mu\text{s}^{-1}$ under highly stabilizing conditions. Using the free energy barriers from this model, the calculated folding rate agrees well with the measured rate extrapolated to zero denaturant, but not at the transition midpoint. Considering other reasonable values for the parameters in the model suggests that the calculated prefactor may be off by a factor of about 3. These calculations also suggest that the folding rate prefactor near water solvent conditions scales linearly with solvent viscosity. This dependence agrees with experimental studies on the other two state proteins^{80–82} and simulations of off-lattice model proteins.⁸³ Nevertheless, these conditions are near a turnover regime for which the internal friction of the polymer chain dominates, ultimately causing the prefactor to saturate upon decreasing solvent viscosity. Thus folding of polymers with other backbones may be quite different.

The discrepancy between the calculated and measured folding rate at the transition midpoint (T_f) suggests that the model underestimates the activation barrier or overestimates prefactor for this stability. Since the model has many simplifications that could influence both the thermodynamics and the dynamics, it is impossible without further investigation to identify which primarily needs improvement. It may be combination of both.

One aspect of the present model is that the native state is quite a bit more flexible than observed in the laboratory, with RMS fluctuations on the order of the distance between adjacent monomers a (the spatial resolution of the the harmonic chain). In comparison, temperature factors from x-ray crystallography correspond to RMS fluctuation of less than 1 \AA ($a/4$). A more realistic non-Gaussian polymer backbone would increase the resolution of the model. Such a non-Gaussian model could still be treated using the variational formalism. Still retaining the Gaussian description of the chain, multibody forces arising from neglected degrees of freedom would also tend to increase the rigidity of the native state and are a likely contributor. Recent off-lattice simulations have shown that effective multibody forces sharpen interfaces between folded and unfolded regions and increase activation barriers.⁸⁴ As discussed in paper I, contiguous sequence approximations^{13–15} also give inherently sharp interfaces and larger activation barriers even for pairwise additive forces since the conformations of the residues are restricted to be either folded or unfolded. (The contiguous sequence approximations with multibody forces would give even higher barriers.) If multibody forces are indeed important, the completely folded or unfolded conformations assumed in these approximations may well be more appropriate than expected from simulations of lattice and off-lattice models with only pairwise forces. Still, the interfaces are probably not as cleanly sharp as envisioned in the contiguous sequence approximations. Using the calculated prefactor at the transition midpoint, $k_0 \approx 8 \mu\text{s}^{-1}$, and the measured folding rate of 100 s^{-1} gives a barrier of approximately $11k_B T_f$. This value is greater than both the activation barrier from the variational theory ($\Delta F^\ddagger \approx 5k_B T_f$) and the double sequence approximation calculated in paper I ($\Delta F^\ddagger \approx 8k_B T_f$). Multibody forces may sharpen the distinction between folded and unfolded regions in the transition state ensemble to give an activation barrier consistent with the prefactor calculation at the transition midpoint.

Alternatively, it is possible that the discrepancy at T_f arises because of neglected dynamical effects that could slow the chain dynamics, reducing the folding rate prefactor. In this paper, the internal friction is interpreted as a correction to account for the activated isomerization reaction of the backbone. As mentioned in the introduction, there are other possible contributors to this friction. For example, non-native interactions are presently neglected in the model. These interactions contribute to the memory friction and therefore the internal viscosity. They can be treated through mode coupling calculations similar to those we have done for a random heteropolymer⁸⁵ (see also Refs. 86–88). A more careful treatment of the solvent may also be necessary for an accurate description of the chain dynamics. Detailed calculations

of the reaction dynamics for isomerization transition in di-alanine (a much less complicated system) suggest the solvent may be an essential component of even the local reaction coordinate.⁷⁶ Furthermore, contact formation dynamics may also be slowed because contact formation excludes the solvent, thereby adding an additional activation barrier and time scale into the dynamics. Supporting this possibility, we note that simulations of minimal protein folding models with desolvation barriers in the pair potential bring simulations into qualitative agreement with pressure studies on folding rates.⁸⁹

Since we expect some of these neglected internal friction contributions to influence protein folding dynamics, we have been careful in this paper to use “bare” parameters (and investigate other reasonable estimates) instead of treating them as fitting parameters in the theory. This is an attempt to avoid missing a signal that such new dynamical considerations are indeed important by using inappropriate effective parameters. The folding rate alone, however, is not enough to determine the accuracy of the calculated prefactor, since the barrier height may not yet be accurately enough determined.

Consequently, these proposed effects are perhaps best studied at this point using more direct experimental probes such as intrachain fluorescence quenching measurements. Fluorescence quenching is a powerful tool to study the average structure of ensembles^{90,91} and has recently been used as a probe for experiments on single molecules that allow distributions to be addressed.^{92,93} Motivated primarily by the relevance to protein folding dynamics, some fluorescence quenching experiments have focused on less complicated systems such as short peptides.^{94–96} These measurements of simple polymer pair dynamics can guide the development of improved theoretical models helpful to understanding protein folding kinetics. This is particularly evident in the present model, since the expression for the effective mobility [Eq. (43)] is formally related to approximations to the intrachain quenching rate in polymers.^{66–68}

Experiments monitoring the intrachain contact formation through fluorescence quenching suggest that the dynamical issues mentioned above play a role in the pair contact dynamics as well. Fitting data to effective models for the pair dynamics, the apparent diffusion coefficient comes out to be 10–20 times smaller than one would expect from the diffusion coefficients of the free probes.^{25,90,91,94,96} In those models, the polymer is treated as a single pair connected by an entropic spring controlling the mean square fluctuation of chain ends⁹⁷ where the effective diffusion coefficient of the pair, $D_{\text{eff}}=2D_0$, is a short time approximation to the full dynamics including all the monomers.⁹⁸ Haas has suggested that the internal friction of the chain may explain the discrepancy between D_{eff} and the fitted value,⁹⁴ but no detailed calculation has been done. This is a problem we wish to return to.⁹⁹ Doi’s arguments suggest that the short time properties may dominate.⁶⁸ We note that the inclusion of internal viscosity from the bond flipping rate as presented here ($\gamma_I/\gamma_0 \approx 10$) reduces the effective short time diffusion coefficient¹⁰⁰ to a value $D_{\text{eff}} \approx 0.5D_0$, narrowing the discrepancy to within a factor of about 2–5. Since the validity of

this short time approximation can be questioned (e.g., the rate depends on the quenching volume),^{68,98} we feel a solution resolution to this puzzle requires a more complete calculation.

The primary experimental technique used to study the structure of the transition state ensemble is the ϕ -value analysis developed by Fersht.¹⁰¹ By comparing the change in the folding rate to the change in stability upon mutation, one can infer the extent to which that site is structured in the transition state ensemble. To interpret these experiments, one assumes that the mutation alters the kinetics through a change in the free energy profile, leaving the prefactor unchanged. Sometimes, however, the kinetics and thermodynamics are seen not to be directly coupled. This may well be a sign of frustration from nonstabilizing native contacts or non-native contacts influencing the kinetics of the transition.¹¹ It is also possible however that these mutations affect the chain dynamics by modifying the internal friction at that residue. It is easy to extend the present calculation to indicate how the internal friction could be exploited to find “dynamical mutants” in which a residue is replaced by one with higher dihedral barriers (i.e., a residue with larger local internal friction) perturbing the prefactor, without altering the stability. Proline isomerization coupled to folding are an example of this, but other non-natural amino acids may be still more promising probes of local reaction coordinates.¹⁰² Mutations of the n-src loop of the SH3 domain have very small effect on the equilibrium constant (preventing the ϕ -value analysis), but still impact the folding rate and thus may be examples of dynamical mutants.¹⁰³ More extreme alterations to the sequence have been necessary such as lengthening the loop length in order to characterize this region of the transition state ensemble.¹⁰⁴ A better understanding of the chain dynamics should help suggest additional and more direct probes of similar situations using dynamical mutants.

ACKNOWLEDGMENT

This work has been supported by NIH Grant No. PHS 2 RO1 GM44557.

APPENDIX

In this appendix, we outline the calculation of the native density correlations $\langle \rho_i(t) \rho_j(0) \rangle_0$ given by Eq. (36).

Since H_0 is harmonic, the Green’s function for the diffusion operator is known and has the form of a multivariable Gaussian containing all the pair correlations.⁶³ The native density correlations only depend on two monomer positions for a fixed i and j , so that Eq. (27) can be simplified by introducing delta functions $\int d\mathbf{r}_1 d\mathbf{r}_2 \delta(\mathbf{r}_1 - \mathbf{r}_i) \delta(\mathbf{r}_2 - \mathbf{r}_j)$ and integrating over all monomer variables $\{\mathbf{r}\}$. This reduces Eq. (27) to a double integral over \mathbf{r}_1 and \mathbf{r}_2 , and the reduced Green’s function contains only the monomer correlations of the monomer pair, as can be expected for a Gaussian process.

It is perhaps easier to make use of this Gaussian property from the beginning, and obtain the same result as using the full Green’s function as described above.

Consider a fixed i and j pair. First we define some notation. We define the following two-component vectors:

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_i(t) \\ \mathbf{r}_j(t') \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \mathbf{s}_i \\ \mathbf{s}_j \end{bmatrix}, \quad \mathbf{R}^N = \begin{bmatrix} \mathbf{r}_i^N \\ \mathbf{r}_j^N \end{bmatrix}, \quad (\text{A1})$$

where t and t' denote two times, and e.g., $\mathbf{s}_i = \langle r_i \rangle_0$ and \mathbf{r}_i^N is the average and native position vector of the i th monomer, respectively. Let $\chi(t-t')^{-1}$ be the symmetric 2×2 matrix of monomer correlations $\chi(t-t')^{-1} = \langle \delta \mathbf{R}(t) \cdot \delta \mathbf{R}(t')^T \rangle_0 / a^2$, with $\delta \mathbf{R}(t) = \mathbf{R}(t) - \mathbf{S}$:

$$\chi(|t-t'|)^{-1} = \begin{bmatrix} G_{ii} & G_{ij}(|t-t'|) \\ G_{ij}(|t-t'|) & G_{jj} \end{bmatrix}. \quad (\text{A2})$$

To construct $\chi(|t-t'|)$, we have denoted the equal time correlation function by the static correlations G_{ii} and G_{jj} , and used symmetry of the correlations $G_{ij}(t) = G_{ji}(t)$.

Since this is a Gaussian process, the correlation function $\langle \rho_i(t) \rho_j(t) \rangle_0$ can be written as

$$\langle \rho_i(t) \rho_j(t') \rangle_0 = \int_{\mathbf{r}_i, \mathbf{r}_j} \rho_i(\mathbf{r}_i) \rho_j(\mathbf{r}_j) \tilde{P}(\delta \mathbf{r}_i, \delta \mathbf{r}_j; |t-t'|), \quad (\text{A3})$$

where

$$\begin{aligned} \tilde{P}(\delta \mathbf{r}_i, \delta \mathbf{r}_j; |t-t'|) &= \left(\frac{3}{2\pi a^2} \right)^3 (\det \chi(|t-t'|))^{3/2} \\ &\times \exp \left[-\frac{3}{2a^2} \delta \mathbf{R}^T \cdot \chi(|t-t'|) \cdot \delta \mathbf{R} \right]. \end{aligned} \quad (\text{A4})$$

In the rest of the derivation, we suppress the time dependence to simplify notation. Substituting Eq. (4) for $\rho(\mathbf{r}_i) \rho(\mathbf{r}_j)$ and shifting the integration variables to $\delta \mathbf{R}$ ($\mathbf{R} = \delta \mathbf{R} + \mathbf{S}$) leads to

$$\begin{aligned} \langle \rho_i(t) \rho_j(t) \rangle_0 &= \left(\frac{3}{2\pi a^2} \right)^3 (\det \chi)^{3/2} \exp \left[-\frac{3\alpha^N}{2a^2} (\mathbf{R}^N - \mathbf{S})^2 \right] \\ &\times \int_{\delta \mathbf{r}_i, \delta \mathbf{r}_j} \exp \left[-\frac{3}{2a^2} \delta \mathbf{R} \cdot \chi' \cdot \delta \mathbf{R} \right. \\ &\left. + \frac{3}{a^2} \mathbf{J} \cdot \delta \mathbf{R} \right], \end{aligned} \quad (\text{A5})$$

where

$$\chi' = \chi + \alpha \mathbf{I}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (\text{A6})$$

$$\mathbf{J} = \mathbf{R}^N - \mathbf{S}. \quad (\text{A7})$$

Performing this integral by completing the square in the exponent gives

$$\begin{aligned} \langle \rho_i(t) \rho_j(t) \rangle_0 &= \left[\frac{\det \chi}{\det \chi'} \right]^{3/2} \exp \left[-\frac{3\alpha^N}{2a^2} \mathbf{J}^2 \right] \\ &\times \exp \left[\frac{3(\alpha^N)^2}{2a^2} \mathbf{J}^T \cdot \chi'^{-1} \cdot \mathbf{J} \right]. \end{aligned} \quad (\text{A8})$$

To simplify this equation, it is convenient to rewrite χ' as

$$\chi' = \chi M, \quad (\text{A9})$$

where

$$\begin{aligned} M &= (I + \alpha^N \chi^{-1}) \\ &= \begin{bmatrix} 1 + \alpha^N G_{ii} & \alpha^N G_{ij}(|t-t'|) \\ \alpha^N G_{ij}(|t-t'|) & 1 + \alpha^N G_{jj} \end{bmatrix}. \end{aligned} \quad (\text{A10})$$

Then we can express χ'^{-1} as

$$\chi'^{-1} = M^{-1} \chi^{-1}. \quad (\text{A11})$$

Now since

$$\mathbf{I} = M^{-1} M \quad (\text{A12})$$

$$= M^{-1} (I + \alpha^N \chi^{-1}) \quad (\text{A13})$$

$$= M^{-1} + \alpha^N M^{-1} \chi^{-1} \quad (\text{A14})$$

$$= M^{-1} + \alpha^N \chi'^{-1}, \quad (\text{A15})$$

we can express χ'^{-1} as

$$\chi'^{-1} = \frac{1}{\alpha^N} (I - M^{-1}). \quad (\text{A16})$$

Consequently, the combined argument of the exponential factors in Eq. (A8) is

$$\frac{3(\alpha^N)^2}{2a^2} \left(\mathbf{J}^T \cdot \chi'^{-1} \cdot \mathbf{J} - \frac{1}{\alpha^N} \mathbf{J}^2 \right) = -\frac{3\alpha^N}{2a^2} \mathbf{J}^T \cdot M^{-1} \cdot \mathbf{J}. \quad (\text{A17})$$

The determinate factor can be simplified using

$$\det \chi' = \det \chi \det M. \quad (\text{A18})$$

Thus, Eq. (A8) becomes

$$\langle \rho_i(t) \rho_j(t) \rangle_0 = (\det M)^{-3/2} \exp \left[-\frac{3\alpha^N}{2a^2} \mathbf{J}^T \cdot M^{-1} \cdot \mathbf{J} \right], \quad (\text{A19})$$

which is Eq. (36).

- ¹M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Oxford University Press, Oxford, 1986).
- ²R. Zwanzig, *J. Chem. Phys.* **60**, 2717 (1974).
- ³M. Bixon and R. Zwanzig, *J. Chem. Phys.* **68**, 1896 (1978).
- ⁴S. E. Jackson, *Folding Des.* **3**, R81 (1998).
- ⁵L. S. Itzhaki, D. E. Otzen, and A. R. Fersht, *J. Mol. Biol.* **254**, 260 (1995).
- ⁶I. Lopez-Hernandez and L. Serrano, *Folding Des.* **1**, 43 (1996).
- ⁷R. E. Burton, G. S. Huang, M. A. Daugherty, T. L. Calderone, and T. G. Oas, *Nat. Struct. Biol.* **4**, 305 (1997).
- ⁸B. Nölting *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 826 (1997).
- ⁹V. P. Grantcharova, D. S. Riddle, J. V. Santiago, and D. Baker, *Nat. Struct. Biol.* **5**, 714 (1998).
- ¹⁰J. C. Martinez, M. T. Pisabarro, and L. Serrano, *J. Mol. Biol.* **5**, 721 (1998).
- ¹¹B. A. Shoemaker, J. Wang, and P. G. Wolynes, *J. Mol. Biol.* **287**, 675 (1999).
- ¹²J. J. Portman, S. Takada, and P. G. Wolynes, *Phys. Rev. Lett.* **81**, 5237 (1998).
- ¹³O. V. Galzitskaya and A. V. Finkelstein, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 112999 (1999).
- ¹⁴E. Alm and D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11305 (1999).
- ¹⁵V. Muñoz and W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 11311 (1999).
- ¹⁶V. I. Abkevich, A. M. Gutin, and E. I. Shakhovich, *Biochemistry* **33**, 100026 (1994).

- 17 J. N. Onuchic, N. D. Socci, Z. A. Luthey-Schulten, and P. G. Wolynes, *Folding Des.* **1**, 441 (1996).
- 18 E. I. Shakhnovich, *Curr. Opin. Struct. Biol.* **7**, 29 (1997).
- 19 A. J. Li and V. Daggett, *J. Mol. Biol.* **257**, 412 (1996).
- 20 D. K. Klimov and D. Thirumalai, *J. Mol. Biol.* **282**, 471 (1998).
- 21 J.-E. Shea, J. N. Onuchic, and C. L. Brooks III, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12512 (1999).
- 22 V. S. Pande and D. S. Rokhsar, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1273 (1999).
- 23 J. D. Bryngelson and P. G. Wolynes, *J. Phys. Chem.* **93**, 6902 (1989).
- 24 Z. Guo and D. Thirumalai, *Biopolymers* **36**, 83 (1995).
- 25 S. J. Hagen, J. Hofrichter, A. Szabo, and W. A. Eaton, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 11615 (1996).
- 26 E. Helfand, *J. Chem. Phys.* **54**, 4651 (1971).
- 27 J. Skolnick and E. Helfand, *J. Chem. Phys.* **72**, 5489 (1980).
- 28 E. Helfand and J. Skolnick, *J. Chem. Phys.* **77**, 5714 (1982).
- 29 G. Allegra and F. Ganazzoli, *J. Chem. Phys.* **74**, 1310 (1981).
- 30 A. Perico, S. Bisio, and C. Cuniberti, *Macromolecules* **17**, 2686 (1984).
- 31 A. Perico and M. Guenza, *J. Chem. Phys.* **84**, 510 (1986).
- 32 A. Perico, F. Ganazzoli, and G. Allegra, *J. Chem. Phys.* **87**, 3677 (1987).
- 33 A. Perico, *Acc. Chem. Res.* **22**, 336 (1989).
- 34 A. Perico, *Biopolymers* **28**, 1527 (1989).
- 35 Y. Hu, G. R. Fleming, K. F. Freed, and A. Perico, *Chem. Phys.* **158**, 395 (1991).
- 36 A. Perico, N. E. Moe, and M. D. Ediger, *J. Chem. Phys.* **108**, 1245 (1998).
- 37 W. Kuhn and H. Kuhn, *Helv. Chim. Acta* **29**, 609 (1946).
- 38 R. Cerf, *J. Polym. Sci.* **23**, 125 (1957).
- 39 C. W. Manke and M. C. Williams, *Macromolecules* **18**, 2045 (1985).
- 40 M. Fixman, *Faraday Discuss. Chem. Soc.* **83**, 199 (1987).
- 41 M. Fixman, *J. Chem. Phys.* **89**, 2442 (1988).
- 42 P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1979).
- 43 S. A. Adelman and K. F. Freed, *J. Chem. Phys.* **67**, 1380 (1977).
- 44 H. A. Kramers, *Physica (Utrecht)* **7**, 284 (1940).
- 45 A. M. Berezhlovskii, E. Pollak, and V. Y. Zitserman, *J. Chem. Phys.* **97**, 2422 (1992).
- 46 J. S. Langer, *Ann. Phys. (N.Y.)* **54**, 258 (1969).
- 47 D. W. Oxtoby, in *Liquids, Freezing and the Glass Transition*, edited by D. Levesque, J. P. Hansen, and J. Zinn-Justin (Elsevier, New York, 1991), pp. 149–191.
- 48 S. Takada and P. G. Wolynes, *Phys. Rev. E* **55**, 4562 (1997).
- 49 R. F. Grote and J. T. Hynes, *J. Chem. Phys.* **73**, 2715 (1980).
- 50 J. S. Langer and L. A. Turski, *Phys. Rev. A* **8**, 3230 (1973).
- 51 D. F. Calef and P. G. Wolynes, *J. Chem. Phys.* **78**, 470 (1983).
- 52 R. Zwanzig, in *Lectures in Theoretical Physics* (Wiley, New York, 1961), Vol. 3, pp. 106–141.
- 53 H. Mori, *Prog. Theor. Phys.* **33**, 423 (1965).
- 54 A. Perico, R. Pratalongo, K. F. Freed, R. W. Pastor, and A. Szabo, *J. Chem. Phys.* **98**, 564 (1993).
- 55 P. Grigolini, *Adv. Chem. Phys.* **62**, 1 (1985).
- 56 Y. Hu *et al.*, *J. Chem. Phys.* **93**, 822 (1990).
- 57 X. Y. Chang and K. F. Freed, *J. Chem. Phys.* **99**, 8016 (1993).
- 58 W. H. Tang, X. Y. Chang, and K. F. Freed, *J. Chem. Phys.* **103**, 9492 (1995).
- 59 K. S. Kostov and K. F. Freed, *J. Chem. Phys.* **106**, 771 (1997).
- 60 A. Perico and R. Pratalongo, *Macromolecules* **30**, 5958 (1997).
- 61 S. Fausti, G. L. Penna, C. Cuniberti, and A. Perico, *Biopolymers* **50**, 613 (1999).
- 62 G. L. Penna, R. Pratalongo, and A. Perico, *Macromolecules* **32**, 506 (1999).
- 63 H. Risken, *The Fokker-Plank Equation*, 2nd ed. (Springer-Verlag, Berlin, 1989).
- 64 H. Yamakawa, *Modern Theory of Polymer Solutions* (Harper & Row, New York, 1971).
- 65 As in paper I, the sequence index begins with the first residue of the fast folding double mutant (G46A/G48A) of λ_{6-85} (Ref. 77). The helices roughly correspond to the sequence indices H1(3–21), H2(27–36), H3(38–47), H4(53–65), H5(72–80) as determined by the DSSP algorithm (Ref. 105) applied to the PDB structure (Ref. 106).
- 66 G. Wilemski and M. Fixman, *J. Chem. Phys.* **60**, 866 (1973).
- 67 G. Wilemski and M. Fixman, *J. Chem. Phys.* **60**, 878 (1973).
- 68 M. Doi, *Chem. Phys.* **9**, 455 (1975).
- 69 G. Allegra, *J. Chem. Phys.* **84**, 5881 (1986).
- 70 E. R. Bazúa and M. C. Williams, *J. Chem. Phys.* **59**, 2858 (1973).
- 71 C. W. Manke and M. C. Williams, *J. Rheol.* **33**, 949 (1989).
- 72 R. Du, V. S. Pande, A. Y. Grossberg, T. Tanaka, and E. S. Shakhnovich, *J. Chem. Phys.* **108**, 334 (1998).
- 73 C. Wagner and T. Kiefhaber, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6716 (1999).
- 74 N. D. Socci, J. N. Onuchic, and P. G. Wolynes, *J. Chem. Phys.* **104**, 5860 (1996).
- 75 T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd ed. (W. H. Freeman, New York, 1993).
- 76 P. G. Bolhuis, C. Dellago, and D. Chandler, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5877 (2000).
- 77 R. E. Burton, G. S. Huang, M. A. Daugherty, P. W. Fullbright, and T. G. Oas, *J. Mol. Biol.* **263**, 311 (1996).
- 78 J. D. Bryngelson, J. N. Onuchic, N. D. Socci, and P. G. Wolynes, *Proteins: Struct., Funct., Genet.* **21**, 167 (1995).
- 79 J. Sabelko, J. Ervin, and M. Gruebele, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 6063 (1999).
- 80 M. Jacob, T. Schindler, J. Balbach, and F. X. Schmid, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 5622 (1997).
- 81 K. W. Plaxco and D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13591 (1998).
- 82 M. Jacob and F. X. Schmid, *Biochemistry* **38**, 13773 (1999).
- 83 D. K. Klimov and D. Thirumalai, *Phys. Rev. Lett.* **79**, 317 (1997).
- 84 M. P. Eastwood and P. G. Wolynes, *J. Chem. Phys.* (in press).
- 85 S. Takada, J. J. Portman, and P. G. Wolynes, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2318 (1997).
- 86 J.-R. Roan and E. I. Shakhnovich, *Phys. Rev. E* **54**, 5340 (1996).
- 87 D. Thirumalai, V. Ashwin, and J. K. Bhattacharjee, *Phys. Rev. Lett.* **77**, 5385 (1996).
- 88 E. G. Timoshenko, Y. A. Kuznetsov, and K. A. Dawson, *Phys. Rev. E* **54**, 4071 (1996).
- 89 N. Hillson, J. N. Onuchic, and A. E. Garcia, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14848 (1999).
- 90 D. S. Gottfried and E. Haas, *Biochemistry* **31**, 12353 (1992).
- 91 E. Haas, *IEEE J. Quantum Electron.* **2**, 1088 (1996).
- 92 Y. Jia *et al.*, *Chem. Phys.* **247**, 69 (1999).
- 93 A. A. Deniz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5179 (2000).
- 94 E. Haas, E. Katchalski-Katzir, and I. Z. Steinberg, *Biopolymers* **17**, 11 (1978).
- 95 O. Bieri *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9597 (1999).
- 96 L. J. Lapidus, W. A. Eaton, and J. Hofrichter, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7220 (1999).
- 97 K. Schulten, Z. Schulten, and A. Szabo, *J. Chem. Phys.* **74**, 4426 (1981).
- 98 R. W. Pastor, R. Zwanzig, and A. Szabo, *J. Chem. Phys.* **105**, 3878 (1996).
- 99 J. J. Portman and P. G. Wolynes (unpublished).
- 100 In the dumbbell model, the normalized correlations of the end to end vector $\mathbf{r}_{n1} = \mathbf{r}_n - \mathbf{r}_1$ are approximated to have a single decay constant $\phi(t) = \exp[-t/\tau_{\text{eff}}]$ where $\tau_{\text{eff}}^{-1} = 3D_{\text{eff}}/\langle r_{n1}^2 \rangle$. The effective diffusion coefficient can be identified by comparing the short time expansions of $\phi(t)$ and the corresponding multiexponential relaxation from Eq. (54), resulting in $D_{\text{eff}} = D_0 \sum_p (Q_{pm} - Q_{p1})^2 / \nu_p$. For no internal friction ($\gamma_i = 0$), $Q = Q^R$ are Rouse modes and $\nu_p = 1$, allowing explicit summation over the modes with the result $D_{\text{eff}} = 2D_0$. In general, D_{eff} depends on the internal friction.
- 101 A. R. Fersht, A. Matouschek, and L. Serrano, *J. Mol. Biol.* **224**, 771 (1992).
- 102 M. W. Nowak *et al.*, *Methods Enzymol.* **293**, 504 (1998).
- 103 D. S. Riddle *et al.*, *Nat. Struct. Biol.* **6**, 1016 (1999).
- 104 V. P. Grantcharova, D. S. Riddle, and D. Baker, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7084 (2000).
- 105 W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
- 106 L. J. Beamer and C. O. Pabo, *J. Mol. Biol.* **227**, 177 (1992).